The diversity of residential electricity demand – a comparative analysis of metered and simulated data

Energy and Buildings 151 (2017) 121–131

Contents lists available at ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

The diversity of residential electricity demand – A comparative analysis of metered and simulated data

José Luis Ramírez-Mendiola*, Philipp Grünewald, Nick Eyre

Environmental Change Institute, University of Oxford, UK

ARTICLE INFO

Article history:
Received 2 November 2016
Received in revised form 22 May 2017
Accepted 10 June 2017
Available online 13 June 2017

Keywords:
Domestic electricity use
Energy demand modelling
Electricity demand load profiles

ABSTRACT

A comparative study between simulated residential electricity demand data and metered data from the UK Household Electricity Survey is presented. For this study, a high-resolution probabilistic model was used to test whether this increasingly widely used modelling approach provides an adequate representation of the statistical characteristics the most comprehensive dataset of metered electricity demand available in the UK. Both the empirical and simulated electricity consumption data have been analysed on an aggregated level, paying special attention to the mean daily load profiles, the distribution of households with respect to the total annual demands, and the distributions of the annual demands of particular appliances. A thorough comparison making use of both qualitative and quantitative methods was made between simulated datasets and it's metered counterparts. Significant discrepancies were found in the distribution of households with respect to both overall electricity consumption and consumption of individual appliances. Parametric estimates of the distributions of metered data were obtained, and the analytic expressions for both the density function and cumulative distribution are given. These can be incorporated into new and existent modelling frameworks, as well as used as tools for further analysis.

1. Introduction

It is a well-known fact that the residential sector is a major contributor to overall electricity consumption in most countries. Residential electricity consumption in the UK accounts for over a third of the national total [5]. Moreover, it is also well known that the domestic sector contributes significantly to peak demand, especially during winter. Over the last decade, however, the potential of this sector to contribute towards reductions in energy consumption and CO_2 emissions has been increasingly recognised. This, in turn, has sparked the interest in this research area.

In particular, there is an increasing interest in understanding the consumption patterns of the residential sector, and how these might change in response to changes in climate and energy prices, and the implementation of new supply-demand balancing strategies and other low-carbon measures.

To this end, several efforts have been made to model electricity demand, seeking to quantify the energy requirements at different levels of analysis. Two fundamentally different approaches divide these modelling efforts: top-down and bottom-up. This terminology refers to the hierarchical level of the data inputs. Top-down models focus on describing the overall trends observed in historical records, with little or no interest in individual end-uses. In contrast, bottom-up models focus on representing the contribution of each end-use towards the aggregate consumption and overall trend. A more detailed description of these two approaches can be found in [19].

Electricity demand modelling has proved very important in both academia and industry. In particular, bottom-up models' ability to incorporate many different factors into the modelling has been key to identifying and understanding the essential elements associated with the production of demand loads. As the transition towards low-carbon energy systems progresses, robust modelling tools become even more relevant. The effectiveness of low-carbon measures such as the implementation of Demand-Side Management (DSM) strategies relies heavily on the extent to which we understand residential electricity consumption. A better understanding of the electricity consumption patterns, and the potential changes these may undergo, would allow us to devise the most suitable measures.

* Corresponding author.
E-mail address: joseluis.ramirezmendiola@ouce.ox.ac.uk
(J.L. Ramírez-Mendiola).

http://dx.doi.org/10.1016/j.enbuild.2017.06.006

As new, relevant data become available, opportunities for further improvement in our models arise. The lack of information about consumer behaviour and appliance usage during the development of models such as the one used in this study (see Section 3) may lead to problems which could not have been revealed with the data available at the time. However, by applying these models in a different context to the one in which they were validated, these issues can now be revealed and explored. Moreover, the new sources of data currently available provide us with an opportunity to re-assess the models' performance. In particular, it is worth verifying that the simulated electricity consumption data produced by the models aggregate such that the statistical characteristics observed in the metered data are well represented, as this is key to a robust bottom-up model.

In the following section brief descriptions of the evolution of residential energy demand modelling and the state of the art are given. The next two sections concern a description of the key elements of the analysis presented, namely the probabilistic demand model used and the sub-metered electricity consumption data. The methodology for the comparative analysis, which encompasses both qualitative and quantitative methods, is discussed in Section 5.1. This is followed by a discussion of the analysis results, where the critical shortcomings identified are highlighted. Finally, a summary of concluding remarks based on the results previously described is given, as well as further details about potential directions of future work.

2. Residential electricity demand modelling

In terms of demand for electricity, the variability across households is high. In the UK, for instance, according to data from the Office of Gas and Electricity Markets (Ofgem), 'Typical Domestic Consumption Values' for electricity range from 2000 to 7200 kWh/year [13]. The results of the most detailed study (to date) of electricity use in English households would appear to indicate that there is an even higher variability, with values ranging from 560 to 14580 kWh/year [16].

There are many factors associated with this variability. The variations in energy consumption associated with differences in dwelling characteristics and the impact of ambient temperature/climate conditions have been extensively studied. However, the impacts of the differences in dwellings' appliance content, and the usage patterns of such appliances have not been studied as much.

Accounting for the impact different combinations of these factors have is critical in representing the range of consumptions currently observed. Consequently, the bottom-up approach to demand modelling has been favoured for this purpose. Methodologies that follow this approach offer the advantage of modelling demand with greater level of detail. Demand estimates can be calculated at the household or even end-use level.

A branch of bottom-up modelling is characterised by a focus on user activity-based simulations of demand patterns. Models falling within this group can be further divided into deterministic and probabilistic models. Deterministic models are based on assumptions about seemingly direct causal links between chosen drivers and expected outcomes. In contrast, probabilistic models make use of stochastic methods for the simulation of electricity consumption patterns. This kind of models make predictions about different outcomes based on the stochasticity of the input data. These models have, therefore, the ability to represent the diversity of energy demand across time and populations.

The model developed by Yao and Steemers [26] is a good example of a deterministic model. The simulation of demand loads is based on fixed, pre-determined household activity patterns characterised mainly by the periods of absence. Appliances are allocated based on national ownership statistics. Their associated demand is calculated based on average ratings and frequency of use estimates. The model is able to produce rough estimates of half-hourly daily load profiles. Another example of a deterministic model is presented in [23].

Recognising the limitations of the deterministic approach, several attempts to capture the variability observed in real households were made. An emerging approach, which continues to attract interest, is based on the use of data from Time-Use Surveys (TUS) for the extraction of characteristic activity patterns. These, in turn, are used in the development of probabilistic models. Page et al. [15] developed a general methodology for the use of Markov chain modelling techniques for the simulation of activity sequences. Several variations of this methodology, supported by the use of country-specific TUS data were implemented. Some of the most commonly cited examples are the models developed by Widén and Wäckelgård [24] and Richardson et al. [18], which are based on Swedish and UK data, respectively. The reader is referred to [20] for a more comprehensive review of models based on this approach.

A few attempts at improving this widely adopted modelling approach have been made. These, however, have focused on the refinement of the simulation of occupancy profiles. To this end, different methods have been applied.

One of these approaches has focused on improving the representation of the duration of the occupancy states through the use of semi-Markov models. In the model presented in [25] the transition probabilities are based on the French TUS, and Weibull distributions are used to model the duration of the occupancy states. In the model presented in [1] the transition probabilities are based on the Belgian TUS and non-parametric estimates are used to model the duration of the occupancy states.

Another approach has focused on the identification of characteristic activity patterns which exhibit statistically significant differences. In the model presented in [4] a Bayesian clustering technique is used for the identification of household-level activity profiles. In the model presented in [1], a hierarchical agglomerative clustering approach is used for identifying characteristic individual occupancy patterns. In both models, the clustering approach is complemented by the development of a Markov model for the identified clusters.

The model presented in [10] is an extension of the model developed by Richardson et al. [18]. In addition to the extension of the model to include thermal demand loads, the simulation of occupancy sequences was refined. The original model was based on two-state occupancy patterns simulations. The refined version is based on the simulation of four-state occupancy patterns. Both models, however, make use of the first-order Markov chain technique for the production of the occupancy sequences. The electricity demand model was left largely unaltered. The only modification being the removal of end-uses associated with thermal demand loads. This original model for electricity demand has found widespread applications in academia and industry, for instance [6,11,3,2,8].

The structure of the model developed by Richardson et al. [18], also known as the CREST model, shows great similarity to that of other, more recent models. There are, however, differences between some of the key elements of the structure. The main difference concerns the approach to simulating occupancy patterns, as discussed in the paragraphs above. A lesser difference concerns the use of country-specific TUS data and statistics on appliance ownership and average ratings. Despite these differences, the approach to converting activity patterns into electricity demand loads has remained essentially the same. Therefore, although this model is based on UK-specific data, the results of the analysis presented in this paper are relevant to any model using a similar approach. In the following section, a more detailed description of the CREST model is presented.

3. The CREST model

In 2010 a new model for residential electricity demand was developed at CREST (Centre for Renewable Energy Systems Technology, Loughborough University). This is a probabilistic model that provides an estimate of the electricity consumption of individual households based on the number of residents, occupancy patterns and dwelling appliance content. Based on the approach developed by Page et al. [15] for the implementation of Markov chain models based on time-use data, Richardson et al. developed a residential building occupancy model [17]. Later on, this occupancy model was combined with the approach developed by Yao and Steemers [26] for generating switch-on events of electric appliances to create a new probabilistic model that has found widespread application in the literature [18]. The model is capable of generating electricity load profiles of individual households as derived from the simulated use of individual appliances.

3.1. Structure

The model is divided into two modules. The first one is responsible for the simulation of dwelling occupancy patterns. The second one is responsible for the conversion of the simulated activity patterns into electricity demand loads. Occupancy profiles are simulated for each household, for each day of the simulated period. These profiles are based upon data from the UK 2000 Time-Use survey [9]. The data from this survey is representative of national time budgets [12]. The version of the model used for the simulations described in this paper features a two-state occupancy model. Individual occupancy profiles are essentially binary sequences of states corresponding to periods of active and inactive occupancy. The active occupancy states also take account of the number of household members currently active.

Each simulated household is allocated a particular set of appliances. Appliances are taken from a pre-defined set of 33 common appliances and randomly allocated based on national ownership statistics. The simulated appliances are configured using statistics that include their mean total annual electricity consumption, average power rating, and average cycle length. This configuration is meant to ensure that the appliances represent some general behaviour or pattern rather than emulate some specific real-life appliance (a specific make, or any other specific characteristics). In this sense, CREST's 'appliances' are rather categories of sorts into which many real-life appliances are expected to fit.

The activation of some of these 'appliances' during the simulations depends on the presence of active occupants in the dwelling. Therefore, the variation in appliance usage is based upon the activity of the members of the household and their number. A demand load profile is generated for each allocated appliance for each day of the simulated period. The loads generated depend on the occupancy profiles previously generated. The total daily load profile is the result of aggregating all of the appliances' daily demand loads.

The variation in electricity consumption between different households is achieved through changes in the household composition, set of appliances, and occupancy patterns of each simulated dwelling. Moreover, a calibration factor provides a means to align the overall average household consumption of the simulated sample to the relevant average annual consumption (e.g. sample group, regional or national average). For the simulations used in the original validation analysis this value was set between 4100 and 4300 kWh; this in order to represent dwellings from a specific region in the UK. The calibration of the model is meant to ensure that the overall average annual consumption of the simulated set of households is around a particular value, which is one of the input parameters for the model.

3.2. Validation

For the purpose of validating the CREST model, household-level electricity consumption data was metered from 22 households in the East Midlands [18]. The households' electricity consumption was monitored for a full year. Thus, the resulting metered dataset contained around eight thousand 1 min resolution daily load profiles. The model was then calibrated against the collected data, and used for simulating the electricity consumption from the 22 monitored households. The simulations runs covered the equivalent of a full year period, and generated data with the same resolution as the metered data.

The simulation output was compared against the original metered data. The validation analysis was based on comparisons of different aspects, including the variation of annual demand between dwellings. One of the aims was to formally assess the extent to which these datasets differ from one another in terms of this variation. To this end, a statistical significance test, namely the Mann-Whitney U test, was used [18]. The purpose of this test is to determine how likely it is that the simulated dataset corresponds to a sample drawn from the same population as the reference sample (metered dataset). One of the advantages of this test is that it can be applied even when the distributions of the data are not known. The resulting measured and simulated annual consumptions are shown below in Fig. 1(A).

In summary, the comparison between metered and simulated datasets appeared to show that there was no significant statistical difference between them, thus validating the model. The CREST model is a good example of the attempts to link activity profiles to electricity consumption via the simulation of appliance usage. For this reason, we were particularly interested in looking at how well represented is this link by the simulations when compared with a larger, highly disaggregated metered dataset. A more detailed description of the data used for this purpose is given below in Section 4.

4. The UK Household Electricity Survey

The UK's Household Electricity Survey (HES) was the result of a study jointly commissioned by Defra, DECC and the Energy Saving Trust [7]. The study had four main objectives:

- To identify the range and quantity of electrical appliances found in the typical home.
- To understand their patterns of use and their impact on peak electricity demand.
- To monitor total electricity consumption of the homes as well as that of individual major appliances in the household.
- To collect user habit data when using the range of appliances present in the households.

For this study 250 owner-occupier households were monitored over the period May 2010 to July 2011. The final report on this study observes that the monitored households were chosen such that the statistical characteristics of the sample would match closely those of the typical socio-economic mix [27]. Only owner-occupiers were recruited for the study. The households, however, were still fairly typical in terms of socio-demographic factors. The average annual electricity consumption across the sample was 4093 kWh/year, which compares very well to the 4115 kWh/year average (as of 2015) across all UK homes [5].

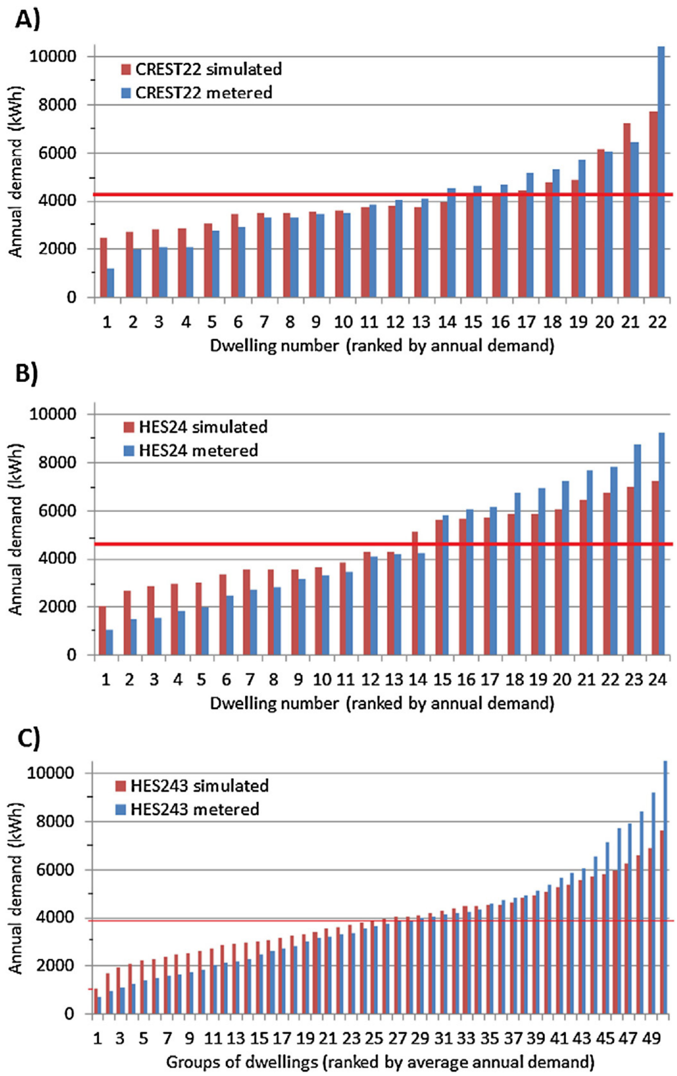

Fig. 1. Annual electricity demands by household, ranked by magnitude: (A) Annual demands of both metered and simulated households in CREST22 dataset (adapted from Ref. [18, Fig. 6]). (B) Annual demands of both metered and simulated households in HES24 dataset. (C) Annual demands of both metered and simulated households in HES243 dataset. Households were grouped in sets of five. The levels observed correspond to the average consumption of each set.

Appliance-level electricity consumption was recorded for each of the 250 households for each day of the monitoring period. For some households total demand load, as read from the mains, was also recorded. The overall length of the monitoring period varies across households. 26 of them were monitored for a full year. The rest were monitored for "one-month-long" periods with effective lengths varying between 20 and 30 days, and covering different parts of the trial period.

The resolution of the metered electricity consumption profiles also varies. For households monitored for a full year, electricity consumption was metered every 10 min. For the rest of the households, it was metered every 2 min. The resulting dataset is a collection of highly detailed electricity consumption profiles, which already has provided very important insights into the way electricity is used in the UK's residential sector. However, the opportunities for exploiting such a rich source are far from exhausted. Further analysis would allow us to have a better understanding of some of the issues associated with residential electricity consumption, such as the scope for demand shifting, estimation of stand-by consumption, the effects of changes in the size and efficiency of appliances, and differences in the way different socio-economic groups use electricity.

The characteristics of this dataset make it ideal for testing current approaches to modelling of residential electricity use. In particular, we are interested in testing whether a model such as CREST could provide us with simulated data that adequately represents the statistical characteristics observed in HES data.

5. Comparison of CREST model's simulations against HES data

The information about the relationship between consumer behaviour and appliance usage was more limited at the time models such as CREST's were developed. The issues caused by such lack of information could not be revealed using the metered electricity consumption data available at the time. However, the arrival of richer datasets allows us to explore these issues now by applying these models in a different context to the one in which they were validated. Now that more comprehensive datasets are available, such as the one provided by HES, this paper re-assesses CREST model's assumptions on appliance usage and their corresponding impacts on daily electricity load profiles and the variability of annual consumption across households.

The interest in testing this model against the HES data resides in great measure in the fact that the group of monitored households is much more diverse than any previous group. Moreover, the appliance-level electricity consumption profiles provide us with a further point of comparison between simulated and metered data.

5.1. Methodology

We wanted to reproduce, to some extent, the original model validation. For the original validation a relatively small sample consisting of 22 households was used (see Section 3.2). Total daily demand for electricity was monitored for a full year. These metered data allowed for comparisons of total annual demands and aggregated load profiles between metered and simulated households. In contrast, HES data offers the opportunity to compare total annual consumption levels, individual appliances' annual consumption and aggregated load profiles of a group of households over 10 times larger.

5.1.1. Data processing

Preliminary analysis of HES data revealed that some of the households monitored during the study are composed of 6+ members. Based on its present configuration, the CREST model can only simulate households of up to 5 members. For this reason and the sake of consistency, the original HES sample was "filtered" so as to be able to make a direct comparison between metered and simulated samples. The final sample resulting from excluding 6+ member households consisted of a total of 243 households, of which 24 had annual records. We will refer to this dataset as HES243. The extent to which this filtering affects the analysis will be discussed in Section 7.

In order to test the simulation outputs against the whole HES dataset, the comparative analysis between simulated and metered data was divided into two parts. Firstly, we focused on the 24 HES households that were monitored for a full year. The corresponding dataset will be referred to as HES24 throughout the rest of the paper. The existence of this sub-set of households with annual records proves important, as it gives us the opportunity to make a closer comparison between the results of the analysis of this HES24 sub-set and the set of 22 households used for the original model validation (See Section 3.2). We will refer to the latter dataset as CREST22.

Table 1
Categories of appliances and the appliances contained in each one, according to the appliances present in both metered and simulated datasets.

Categories	Appliances contained	
	CREST	HES
Cooking	Hob, oven, microwave, kettle	cooker1, cooker2, hob1, hob2, oven1, oven2, hob+oven, microwave1, microwave2, kettle1, kettle2
Washing	Washing machine, Tumble dryer, Dish-washer	Washing machine, washer dryer, Tumble dryer, Dish washer
TV watching	TV1, TV2, TV3	TV1, TV2, TV3, CRT TV, LCD TV, plasma TV, Audio-visual site
ICT related	PC	Desktop PC 1, Desktop PC 2, Computer site, Laptop 1, Laptop 2

Secondly, we then extended the analysis to the rest of the HES sample so as to be able to assess the model's ability to represent the statistical characteristics observed in a larger, more comprehensive metered dataset. However, the sub-set of 219 households for which only monthly records were present had to be processed in a different way to the set of households with full year records.

Total household annual consumptions are not explicitly included in the HES data. This information had to be extracted from the records available for each household. Most households had monthly records only, so these had to be used for producing an estimate of the household's total annual consumption. For each household with monthly records only the available data was used for calculating total daily consumptions for each day in the monitoring period. Then, based on the distribution of total daily consumptions, the total daily consumptions for the missing days needed to complete a full year were generated stochastically to add some variation. The generated daily consumptions were then grouped into the months corresponding to the same monitoring period as that of households with annual records, and the corresponding monthly seasonal factor was applied. Finally, all daily loads were added up thus providing an estimate of the total annual consumption of that particular household. The case of the 24 households with annual records was more straightforward as we only had to add up the consumptions recorded throughout the year. In addition to the total annual consumption, a mean daily load profile for the whole HES sample was calculated based on all available data.

For the simulation of electricity consumption data the CREST model was used. A full version of the CREST model is publicly available in the form of an Excel spreadsheet model (See [18]). However, for the purpose of simulating the electricity consumption of the same number of households as in the HES sample, the CREST model was re-implemented in Python in order to allow for faster, multiple automated runs. Features, functionality and results of Python and Excel versions of the model are equivalent. We used this Python implementation for generating a year's worth of artificial electricity consumption data for 243 households. Each simulated household profile was matched to a metered counterpart in the HES243 dataset. Since it is possible to specify both the day and the month when running simulations, the simulated period was also matched to the monitoring period of the corresponding metered households. Total annual consumptions corresponding to each household were calculated, as well as the mean daily load profile for the whole simulated HES243 sample. For the first part of the analysis, which focused on the HES24 dataset, simulations were run based on the specifications of this sub-set of households only.

As the electricity consumption of individual appliances is available from both metered and simulated data, there was an interest in looking at how the estimates of the annual consumptions of the simulated appliances are distributed and how these distributions compare to those of the corresponding metered appliances. To this end, we also extracted the total annual consumptions of individual appliances for the households in the HES243 dataset.

As observed in Section 3.1, the CREST model generates artificial electricity consumption data based on the simulation of appliance usage for a pre-defined set of appliances. Among the model's pre-defined set, there are 'appliances' which are meant to represent the consumption of the appliances listed in Table 1. This pre-defined set includes some other 'appliances' which represent the electricity consumption of some other devices with more predictable consumption patterns (e.g. cold appliances). However, we focus the analysis on the appliances listed in Table 1, as the electricity loads generated by the use of these appliances are the ones with direct links to households' activities. We compare the simulated demand of these appliances with the data from their corresponding metered counterparts.

The appliances of interest were grouped based on the activities they are associated with: cooking, washing, TV watching, and ICT related activities. Both CREST and HES appliances are grouped in these more general categories, as shown in Table 1. In most cases it is possible to make a one-to-one comparison between HES and CREST appliances in spite of the differences that may exist between them.

In order to obtain the distributions of the simulated total annual consumption of individual appliances, it was necessary to run a new set of simulations. Based on the original model configuration, appliances are allocated to the simulated households on a random basis. Therefore, the original model implementation had to be modified so as to ensure that the appliances of interest were allocated to the relevant households. The goal was to match the appliance configuration of the simulated households to that of the metered counterparts.

HES records show which appliances are owned by which households. Based on this information, households were grouped accordingly. Using the modified model implementation, simulation runs corresponding to one year long periods were carried out.

5.1.2. Comparative analysis

In order to assess whether the simulated data is consistent with the metered one, both qualitative and quantitative methods were used.

The qualitative comparison between datasets focused on the Empirical Cumulative Distribution Function (ECDF) of the different datasets. The ECDF is a formal direct estimate of the Cumulative Distribution Function[1] from which simple statistical properties can be derived. Moreover, the ECDF properties form the basis of various statistical significance tests.

The Probability Density Function (PDF) of the total annual consumptions was obtained as well, as it provides a better graphical representation of some of the statistical characteristics of the metered HES243 dataset. In order to obtain a reliable representation of the PDFs, the Kernel Density Estimation (KDE) method was used. Kernel Density estimation is often used as a data smoothing technique. However, KDE is not simply a smoothing method. This method offers a way of obtaining an accurate non-parametric estimate of the PDF of a given dataset. This estimate is in fact the best possible estimate of the distribution of the original population

[1] Cumulative distribution function (CDF) – It is a function whose value is the probability that a corresponding continuous random variable has a value less than or equal to the argument of the function.

from which the sample was drawn. It can be shown that the kernel density estimate converges to the real distribution as the sample size increases, and it's convergence rate is the highest among all estimates [22].

As observed above, some statistical significance tests are based on the properties of ECDFs. This leads naturally to the quantitative part of this comparative analysis, which focuses on the use of such tests. For the quantitative comparison we will use two tests, namely, the Kolmogorov–Smirnov test and the Mann-Whitney *U* test. Both tests are non-parametric, which means they do not make assumptions about the distribution of the data. Both tests can be used to compare two unpaired datasets, both focus on comparing the statistical characteristics of the distributions of the two datasets being compared, and both are robust to the presence of outliers. However, the two tests work in a very different way.

The Kolmogorov–Smirnov (KS) test compares the cumulative distribution of the two data sets in question, and calculates a *p*-value that depends on the largest difference between the two distributions. The test is sensitive to any differences in the two distributions. Significant differences in shape, spread or median values will result in a small *p*-value. The KS test can also be used to test whether an empirical distribution is consistent with an ideal distribution, and therefore, it is commonly used as a test of goodness of fit, or for testing for normality. In essence, this test answers the question: If the two samples were randomly drawn from different parent populations, how likely it is that the two parent populations have the same statistical distribution?

The Mann–Whitney (MW) test compares two datasets by ranking the elements in the dataset from smallest to largest, and calculating the average of the rank scores of the elements in the two datasets. A *p*-value is calculated which depends on the difference between the average ranks of the two datasets. Compared to the KS test, the MW test is mostly sensitive to changes in the median value. The MW test is closely related to the *t*-test. However, the MW test is more widely applicable than other more popular tests such as the *t*-test, as it does not require the assumption of normally distributed data, and it is much more robust to the presence of outliers. In essence, the MW test answers the question: If both samples come from the same population, how likely it is that random sampling would result in the differences observed in this particular case?

As we are interested in determining whether the simulated data generated by the model accurately capture the statistical characteristics of the metered data, we use these test to obtain a quantitative measure of how closely related the two datasets are. In other words, we use these tests to assess whether both datasets are consistent enough with statistical samples drawn from the same population, or whether the two datasets are consistent with samples drawn from populations which follow the same statistical distribution. The MW test provides us with a quantitative measure of how likely the former situation is, whereas the KS test provides us with a measure of how likely it is the latter.

The MW test was also used in the original validation of the CREST model (see Section 3.2). The positive results of the test were one of the arguments in favour of the validation of the model. We take this as an opportunity to show the effect that sample size may have when using this kind of test by applying the test to the datasets HES24 and HES243 separately and comparing the results.

6. Analysis results

6.1. Variations in annual demand levels

The high variability observed in annual demand levels across the monitored households clearly reflects the fact that different households mean different needs, lifestyles and consumption patterns. In this context, when we talk about variability, we refer to the natural variation that is observed in the demand levels as a result of the differences in the underlying causes of said demand.

Fig. 1 shows both simulated and metered annual demands, obtained from the different datasets. Both the variability of annual demand among households and the differences between the metered and simulated levels can be appreciated, as well as how they compare across datasets. This figure starts to show that the empirical data has greater diversity than the model simulations would suggest. The red horizontal lines indicate the overall average annual consumption of the corresponding metered dataset. Only one line per section is needed because, as observed in Section 3, an adequate calibration of the CREST model ensures that the overall average annual consumption of the simulated sample is around a specified value. The corresponding values are listed in Table 2 which shows, in numbers, how the variability of both simulated and metered data compares across the different datasets.

The mean annual electricity demands per dwelling calculated from metered and simulated data for each of the analysed datasets differ by around 1%. The fact that this difference is small only shows that the model was appropriately calibrated. Despite this, it is observed that the differences in standard deviations are considerably larger. For both HES24 and HES243 datasets the standard deviation of the metered annual consumptions is about 40% higher than the standard deviation of the simulated ones.

The results of the first part of the analysis, concerning the HES24 dataset, already appeared to suggest that the variation in annual demand levels was being under-represented. However, the simulated HES24 dataset used for this study appears to be consistent with the simulated CREST22 dataset used in the original validation analysis[2] (See Table 2).

The analysis of CREST22 datasets showed that the standard deviation of the metered annual consumptions was about 30% higher than the standard deviation of the simulated ones. Despite this seemingly large difference, the result of applying a statistical significance test would appear to indicate that the datasets were not significantly different (See Section 3.2).

6.2. Empirical cumulative distributions and significance tests

The HES24 dataset is very similar in size to the CREST22 dataset used for the original validation. As part of the original validation analysis, a Mann-Whitney (MW) test with a 5% level of significance was used to compare metered and simulated CREST22 datasets. Based on the test results, it was concluded that there was no significant difference between them (Section 3.2). In addition to verifying this, and in order to further test how well represented was the metered data by the simulations in the original analysis, we extracted the annual consumption data from the original validation results and performed a Kolmogorov–Smirnov (KS) test on these datasets. The results of this test also supported the hypothesis that the two datasets are consistent with each other.

The same tests were then applied to metered and simulated HES24 datasets. With a *p*-value of 0.38, the MW test with 5% significance level suggests that it is likely that these two samples were drawn from the same population. Moreover, with a *p*-value of 0.44, the KS test with 5% significance level suggests that it is likely that these two samples come from populations equally distributed. The results of this part of the analysis seem to be well in agreement with those of the original validation, thus confirming that the model produces reasonable estimates when dealing with small samples (∼20 households).

[2] CREST22 data was taken from [18].

Table 2
Mean values and standard deviations of annual electricity consumptions datasets.

	CREST22		HES24		HES243	
	Metered	Simulated	Metered	Simulated	Metered	Simulated
Mean annual demand (kWh/y)	4172	4124	4624.6	4629.7	3944.5	3942.9
Standard deviation (kWh/y)	1943	1372	2510	1545	2348	1405
Coefficient of variation	0.47	0.33	0.54	0.33	0.60	0.36

The second part of the analysis, which focused on the HES243 datasets, was aimed at determining whether the model simulations produce a reasonable representation of the statistical characteristics of a larger, more diverse set of households. Both MW and KS test were used again to provide a quantitative measure of the differences between metered and simulated datasets. On this occasion, however, the tests results revealed the existence of important differences in the statistical characteristics of both datasets. With a p-value of 0.049, the MW test with 5% significance level now suggests that it is unlikely that the metered and simulated HES243 samples come from the same population. In a similar manner, with a p-value of 0.001, the KS test with 5% significance level suggests that it is very unlikely that these two samples come from populations with the same statistical distribution. Moreover, by using the KS test to test for normality, we found that the simulated HES243 dataset appears to be consistent with normally distributed data (p-value of 0.81). The metered HES243 sample is not. These results are best appreciated graphically. As Fig. 2 shows, the simulated HES243 dataset is well in agreement with its corresponding normal fit (dashed line).

Probability Density Functions corresponding to each set of annual consumptions can also be appreciated in Fig. 2. The Kernel Density Estimate of the distribution of metered annual consumptions revealed the existence of a more complex underlying distribution. Based on this estimate, three clusters can be readily identified. What this shows is that the distribution of households with respect to total annual consumption is concentrated in three clusters, centred around the values (modes): 1.58 MWh, 3.92 MWh and 7.85 MWh. Based on this information, and on the same principles behind the Kernel Density Estimation, we produced a parametric estimate for the empirical PDF observed in Fig. 2. This estimate can be expressed by the following trimodal Gaussian mixture:

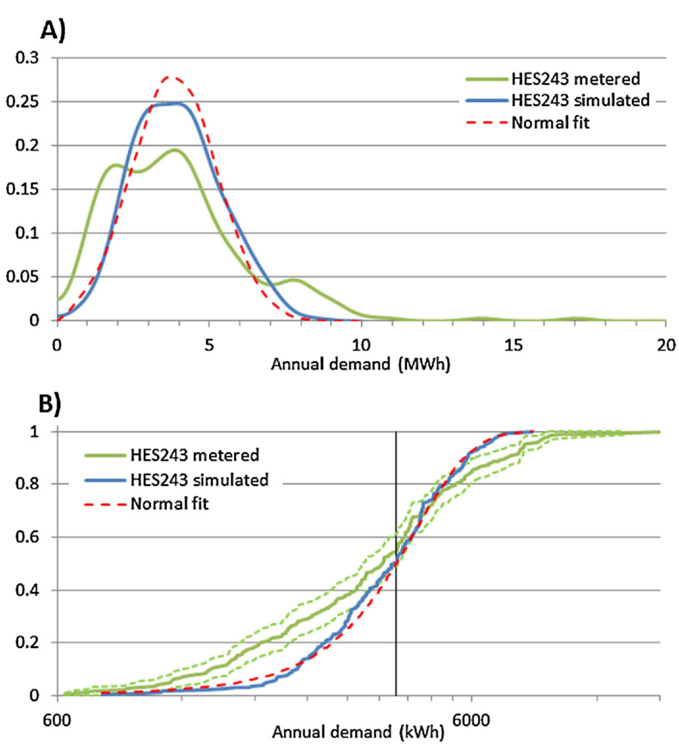

Fig. 2. Distribution of total annual consumptions: (A) Probability Density Functions of metered and simulated HES243 datasets. (B) Empirical Cumulative Distribution of metered and simulated data. The dotted lines around the ECDF of the metered HES243 dataset are the 95% confidence bounds.

Table 3
Trimodal Gaussian mixture model's parameters.

	$\mathcal{N}_1(\mu_1, \sigma_1)$	$\mathcal{N}_2(\mu_2, \sigma_2)$	$\mathcal{N}_3(\mu_3, \sigma_3)$
C_i	0.261	0.622	0.115
μ_i	1.575	3.92	7.85
σ_i	0.78	1.28	1.0

$$f(x) \sim C_1\mathcal{N}_1(x \mid \mu_1, \sigma_1) + C_2\mathcal{N}_2(x \mid \mu_2, \sigma_2) + C_3\mathcal{N}_3(x \mid \mu_3, \sigma_3) \quad (1)$$

where $\mathcal{N}_i(x \mid \mu_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2}$ is the general normal distribution with mean μ_i and standard deviation σ_i, and C_i is a proportionality constant.

A summary of the parameters of this estimate and their corresponding values can be found in Table 3. Fig. 3(A) shows how this trimodal Gaussian model compares to the empirical PDF.

The ability to produce a parametric estimate for the PDF means that it is also possible to find an expression ($F(x)$) for the CDF, as these two functions are related analytically by the equation: $F(x) = \int_{-\infty}^{x} f(t)dt$. Therefore, by integrating $f(x)$ we obtain that the CDF is given by:

$$F(x) = \sum_{i=1}^{3} \frac{C_i}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu_i}{\sigma_i\sqrt{2}}\right)\right] \quad (2)$$

where erf(x) is the error function, and C_i, μ_i and σ_i are the same parameters used in Eq. (1) (see Table 3). The way this estimate compares to the ECDF is shown in Fig. 3(B).

In order to gain a better insight into the relationship between household size and total annual consumption, we looked at the sub-groups of households characterised by these features in two different ways. Firstly, we grouped the households according to size, and used the KDE method to obtain an estimate of the empirical PDF corresponding to each group. These distributions can be observed in Fig. 4(A). The number of households with 5 members is not large enough as to provide a reliable estimate of the overall distribution of annual consumptions for this sub-group. Therefore, this distribution was omitted. Secondly, based on the parametric estimate we obtained for the overall distribution, we grouped the households according to the three observed clusters. The criteria for the allocation of households to the different clusters was whether their annual consumption falls within the interquartile range of the cluster in question. Households in each cluster were further grouped according to size. The percentage composition of each cluster is shown in Fig. 4(B).

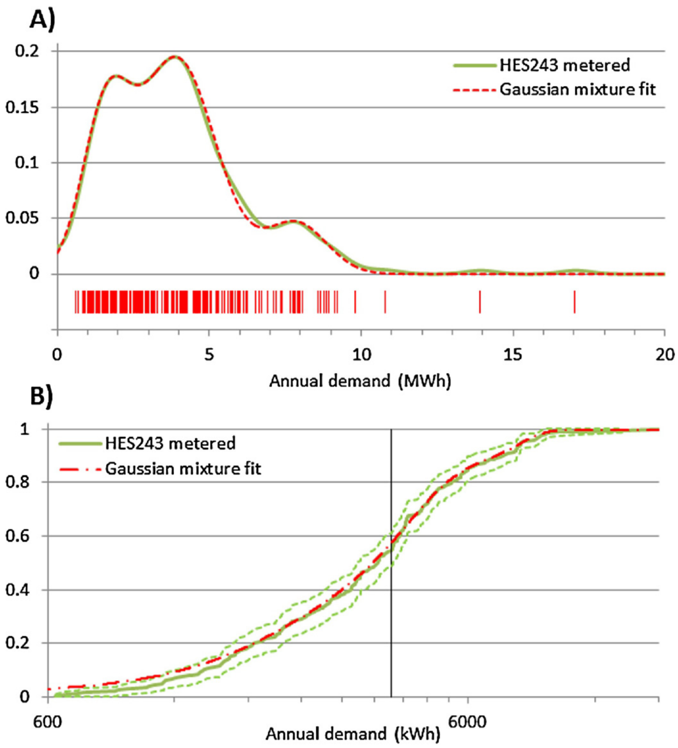

Fig. 3. Total annual consumptions distributions: (A) Probability Density Function of metered HES243 dataset and trimodal Gaussian mixture fit. Each vertical bar at the bottom of the plot represents a data point. (B) Empirical Cumulative Distributions of metered annual consumptions and trimodal Gaussian mixture fit. The dotted lines around the ECDF of the metered HES243 dataset are the 95% confidence bounds.

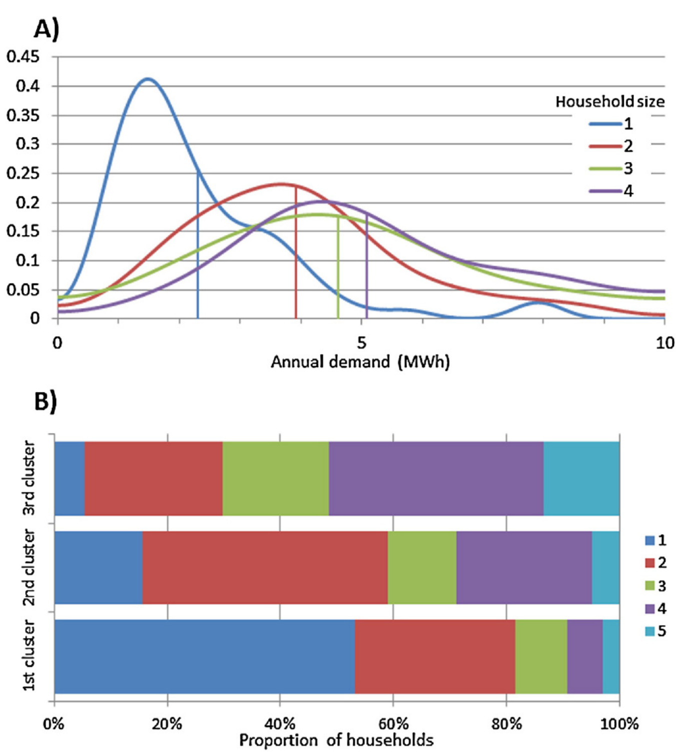

Fig. 4. Distribution of households according to size and annual consumption range: (A) EPDFs corresponding to households grouped by size. The vertical line in each distribution indicates the mean value. (B) Percentage composition of household sizes for each cluster.

Table 4
Summary of parameters of the Weibull distributions fitted to the annual consumptions ECDFs of individual appliances.

	Weibull fit parameters	
	Scale	Shape
Main cooking	356.54	1.26
Cooker	344.72	1.28
Oven	248.34	1.04
Microwave	57.66	1.19
Toaster	21.90	0.94
Kettle	188.20	1.82
Washing machine	164.85	1.20
Tumble dryer	367.87	0.96
Washer dryer	262.81	1.06
Dishwasher	301.57	1.47
TV1	174.90	0.86
TV2	109.22	0.69
Laptop	24.34	0.87
Desktop	156.37	0.89
Computer site	213.01	0.98

6.3. Individual appliances: total annual consumptions

We looked at the appliance content of the HES243 households and, based on the list of appliances of interest presented in Section 5.1, we grouped the households that contained each one of those appliances. We then calculated the ECDF of both simulated and metered data for each one of those groups. Examples of the obtained distributions are shown in Fig. 5.

The results of this analysis made evident that the variability observed in the annual consumption of individual appliances is completely misrepresented by the simulated data. In some cases (see Fig. 5(A) and (D)) the average annual consumption of both datasets is well in agreement. In some others (Fig. 5(B) and (C)), the differences between the average annual consumptions are evident as well.

In general, it was observed that the empirical distributions of individual appliances' annual consumptions were skewed. Log-normal and Weibull models were tested for possible fits, as these models are particularly good at representing skewed data.[3] The analysis revealed that the best fits were provided by Weibull models. The summary of parameters of the distributions fitted to the ECDFs of the annual consumptions of individual appliances is presented in Table 4.

6.4. Mean daily load profile general features

Mean daily load profiles were obtained from both metered and simulated HES243 datasets. These can be observed in Fig. 6, along with the typical UK profile. This typical UK load profile corresponds to the half hour load profile for the average Profile Class 1 customers[4] [21], and is presented for comparison purposes only.

Both the metered and simulated mean daily load profiles have, generally speaking, similar shapes to that of the typical profile. However, when the profiles are compared in more detail, marked contrasts are observed:

[3] A more empirical argument for the use of these models is rooted in the fact that they are generally applicable for modelling size/magnitude distributions (e.g. particle size, wind speed).

[4] Profile Class 1 customers are domestic customers supplied on unrestricted tariffs, with a maximum demand below 100 kW as measured by metering systems containing only one meter register. There are eight profile classes, and for each profile class a sample group of electricity supply customers is randomly drawn from the population of electricity supply market customers. These samples are designed to provide an accurate estimate of the load pattern for each class of customers for use in electricity settlement. Consumption data is obtained by either installing half-hourly meters at the sites or getting half-hourly consumption data directly from suppliers.

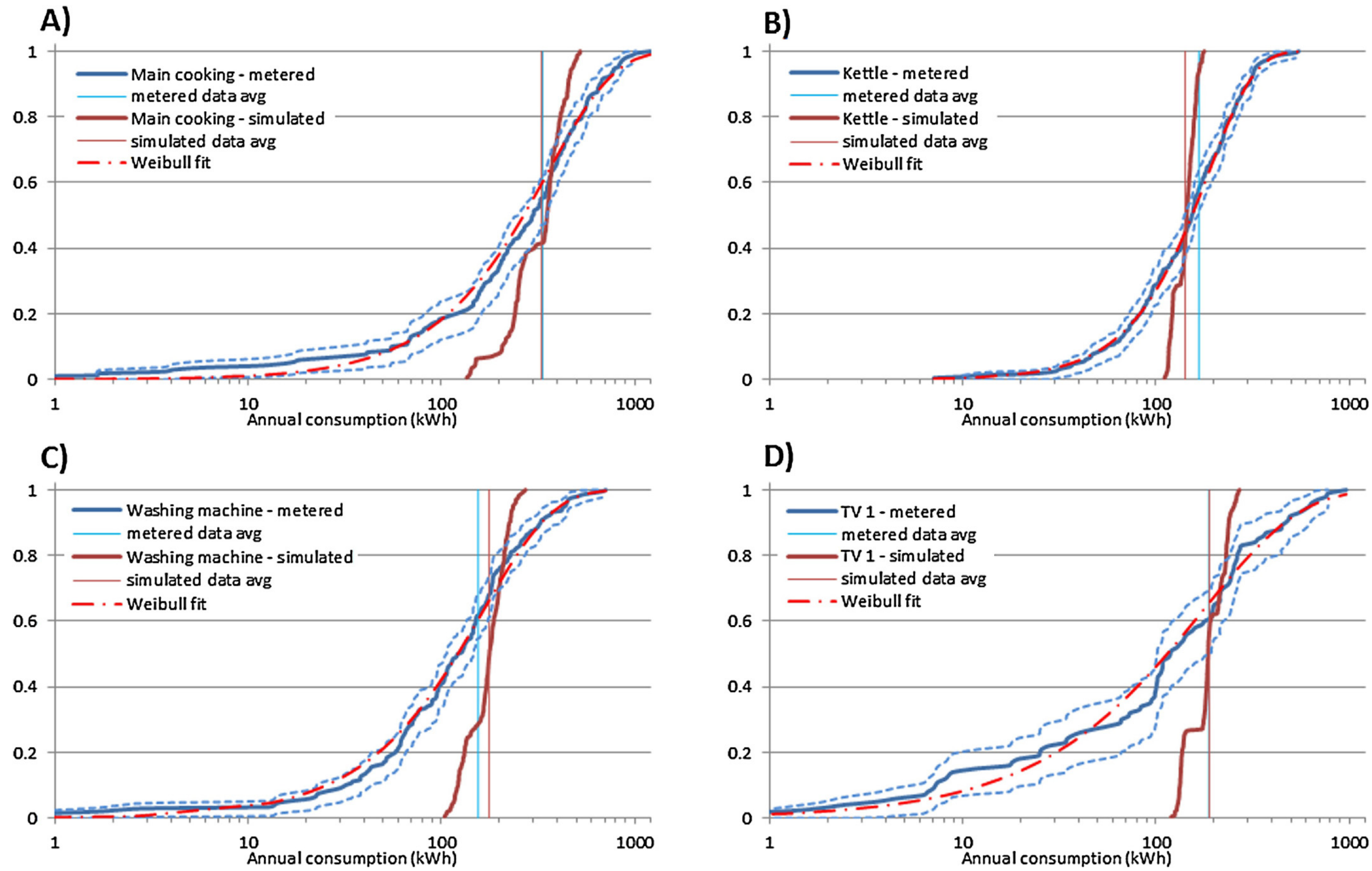

Fig. 5. Annual consumptions ECDFs of individual appliances: (A) Main cooking appliances: cooker, or equivalently, hob + oven. Present in 70% of households. (B) ECDFs of simulated and monitored kettles. Present in 94% of households. (C) ECDFs of simulated and monitored washing machines. Present in 83% of households. (D) ECDFs of simulated and monitored main TVs. Present in 98% of households.

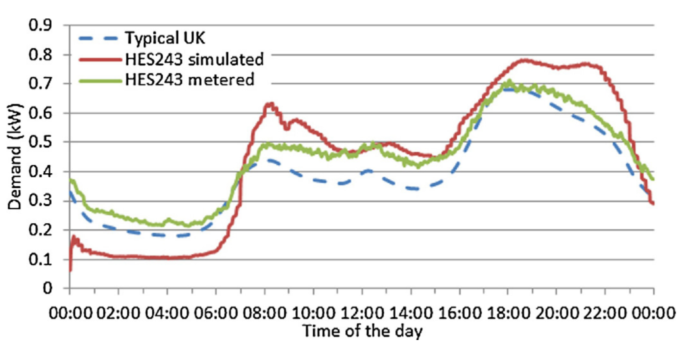

Fig. 6. Mean daily load profiles corresponding to metered and simulated HES243 datasets. Typical UK profile corresponds to the load profile of the average class 1 (unrestricted domestic tariff) customer.

- There are significant differences between the values of the lowest and highest demands observed in the two profiles.
- The simulated profile appears to overestimate the duration of the peak period.
- The variations observed in the CREST profile, such as the transitions between periods of low/high demand and periods of high/low demand, correspond to more abrupt changes than those observed in the other two profiles.

Further analysis of the characteristics and differences of these profiles may be desirable, and indeed, necessary. In particular, it might be interesting to pay a closer look at how the activity profiles and the use of appliances are linked, and how well represented the duration of certain activities is. However, a more detailed analysis of these profiles falls beyond the scope of this paper. The prospects for a more detailed study on the intra-day variability of demand are left to be discussed in the next section.

7. Discussion and future work

The primary aim of the analysis presented in this paper was to determine whether a probabilistic model based on currently used approaches is capable of providing a good representation of the variability observed in the electricity consumption levels of a highly detailed dataset. The choice of the CREST model for this study was due in great measure to the fact that this model seeks to emulate the variability of household electricity use [18]. According to the way in which the model was constructed, the simulated appliances are configured such that their annual consumptions result in figures typical of an average household. Moreover, the overall average annual consumption of the simulated households was also expected to be in agreement with the national figures. The model calibration involves the scaling of the simulated sample's annual consumptions. This is meant to ensure that the simulated average annual consumption matches the average annual consumption of the real-life counterpart (see Section 3).

The high level of detail offered by HES data justifies its use for the analysis. The appliance-level metered electricity consumption records make it ideal for a comparison against the simulated appliance consumption data. Due to the technical limitations of the model discussed in Section 6, the original HES dataset had to be filtered so as to allow for a consistent analysis. The filtering consisted in removing the households with more than five members from the sample. However, we believe this filtering had little impact on the overall statistical characteristics of the dataset and the analysis results. The proportions of households in the HES sample corresponding to household sizes ranging from 1 to 4 are well in agreement with the proportions observed at the national level [14]. If we extrapolate this, by assuming that the proportion of households with five and 6+ residents are equally well represented in the sample, this would mean that households with 6 or more members account for just over 2% of households in the UK. A sense of the scale of the effect of removing said households can be given by comparing the overall average annual consumptions; the average annual consumption across the whole 250 households HES sample is just over 3% higher than the average across the HES243 dataset.

Great part of the analysis presented in this paper focused on the distribution of households with respect to their annual consumptions. The analysis of the annual consumption data extracted from the HES dataset revealed the existence of a complex trimodal distribution (see Section 6.2). Further analysis revealed that the underlying distribution appears to be very well estimated by a trimodal Gaussian mixture with analytical expression given by Eq. (1). Each of the three clusters identified in the empirical distribution is consistent with a normal distribution. The resultant composite distribution, however, does not exhibit normal-like features. A breakdown into sub-groups of households characterised by size revealed that: (1) Over 50% of the households concentrated around the first cluster are one-person households. The first cluster is also the largest one, which means that the majority of the one-person households are found in this consumption range. (2) Over 40% of the households concentrated around the second cluster correspond to two-people households, making this sub-group the predominant one. (3) Over 33% of the households concentrated around the third cluster are 4-people households. The second biggest group corresponds to two-people households (see Fig. 4). It appears to be a linear correlation between the household size of the predominant group and the mean value of the corresponding cluster. However, further analysis is needed in order to determine whether this is a general trend or just an interesting coincidence.

The distributions of the simulated annual consumptions were found to be consistently normal. It would appear that the current simplifying modelling assumptions are causing simulation outputs consistent with normally distributed data. These assumptions would appear to lead to reasonable first-order estimates as misrepresentation issues appear less severe when simulations are restricted to small sets of households. However, the differences between simulated and metered data become evident when simulating larger samples.

Modelling assumptions concerning the conversion of activity profiles into electricity consumption patterns have remained essentially the same over the last decade. It is therefore necessary to reassess these assumptions. Assumptions leading to normally distributed data will need to be modified or replaced entirely. If replacing these assumptions in their entirety proves overly complicated, then the implementation of mixture models based on normal components might be a good alternative.

In addition to the differences in the overall shape of the distributions, it was observed that the variability in the levels of annual consumption is consistently misrepresented in the simulated data. Regardless of sample size, it was found that the level of variability observed in the metered data is consistent across independent datasets. Moreover, the variability observed in the metered data is about 20% higher than in the simulated data (see Table 2). The HES sample is still arguably small, statistically speaking. However, we presume it provides a better insight into the kind of distribution and variability that would be observed in larger samples. Given the method used for estimating the overall distribution, it is also very likely that the statistical characteristics observed in the estimate are close to those of the true population distribution. It is important that a refined representation of the characteristics of the distributions observed in Fig. 3 is achieved, as this is key to a robust approach to residential electricity demand modelling.

The analysis of the distribution of metered annual consumptions generated many interesting results, which include analytic expressions for the Probability Density Function (PDF) and the cumulative distribution function (CDF) (see Section 6.2). Knowing the analytical expression for the cumulative distribution allows for the use of methods such as the inverse transform sampling. Given a cumulative distribution function, this method allows for the generation of samples consistent with data drawn from said distribution. A better idea of the statistical characteristics of the distribution of annual consumptions gives us the opportunity to fine tune demand models. The improvements derived from this would be reflected in simulated datasets that are more in agreement with the kind of distribution and levels of variability observed across real-life households.

Total annual consumptions from individual appliances were used as a further point of comparison between simulated and metered datasets. Based on each dataset, distributions of total annual consumptions were obtained for the appliances listed in Table 1. The choice of the appliances of interest was based on the reasonable expectation that their demand loads would be linked to households activities. In the CREST model's simulations, a calibration factor is calculated for each appliance. According to the model documentation, this factor is adjusted so that over a large number of runs the average annual consumption of the appliance would match typical levels [18]. In some cases it was found that the average annual consumption of the simulated appliances was well in agreement with that of their metered counterparts (see Fig. 5(A) and (D)). In most cases, however, discrepancies were found. In general, the simulated values are consistently higher than those obtained from the metered data. In terms of the distribution, similar problems to those of observed in the distributions of total household annual consumptions were found. The variability observed in the metered data is consistently under-represented by the simulations' output. Moreover, the annual consumptions of individual simulated appliances present skewed distributions. These distributions were found to be consistent with Weibull distributions. A summary of the parameters for the fitted models is presented in Table 4.

A comparison was also made between metered and simulated mean daily load profiles extracted from the corresponding HES243 datasets. Important differences in the overall features of both profiles were identified. Further in-depth analysis is needed in order identify the underlying causes of these issues. However, a more detailed characterisation of said differences falls beyond the scope of this paper. In particular, it would be interesting to investigate in more detail the relationship between household activity patterns, appliance usage and the intra-day variations in demand levels. This will be the subject of future work.

8. Conclusions

In this paper, a comparative analysis of the statistical characteristics of metered and stochastically simulated electricity consumption datasets was presented. Metered household- and appliance-level electricity consumptions were extracted from the UK's Household Electricity Survey dataset (see Section 4). Corresponding electricity consumptions were simulated using a widely implemented probabilistic model based on the UK residential sector (see Section 3). The model was configured such that the simulated households matched the relevant features of the metered counterparts. The analysis of the differences between datasets allowed us to provide a measure of how well represented are the observed features by probabilistic models based on currently used approaches and to identify key shortcomings.

In particular, we compared the way simulated and metered households and individual appliances are distributed with respect to total annual consumptions. An element of the assessment of the model's performance concerns how well represented are these distributions by the simulated output. Based on the analysis presented in this paper we can conclude that current probabilistic models fail to capture some of the key characteristics observed in the metered datasets (see Section 6.2). Significant discrepancies were found when the corresponding distributions were compared (see Figs. 2 and 5). In particular, it was found that metered household annual consumptions follow a complex distribution from which three clusters can be identified. The distributions observed in the simulated data are much simpler. Current underlying assumptions about electricity use appear to lead to normally distributed data with consistently limited variability. Therefore, these assumptions need to be rectified so as to prevent the misrepresentation of the statistical characteristics observed in metered data.

The diversity of real-life households is reflected by the complexity of the distribution of total annual consumptions. A more complex distribution does not necessarily mean over-complicated assumptions need to be used in the models. We propose a trimodal Gaussian mixture model which provides a very good estimate of the observed distribution. Analytical expressions for both the density function and cumulative distribution are given. In a similar manner, the distributions of annual consumptions of individual appliances were found to be well represented by Weibull models. As this paper demonstrates, it is possible to use estimates that capture reasonably well the complexity of the observed distribution while preserving models' simplicity.

The analysis presented shows that residential demand for electricity is more diverse than hitherto assumed. A better idea of how households and individual appliances are distributed with respect to annual consumption provides us with opportunities for further refinement of current probabilistic models. In addition, the results further support the fact that current and future modelling approaches would benefit greatly from the use of larger, highly resolved metered datasets.

Acknowledgements

The authors gratefully acknowledge funding from: the Consejo Nacional de Ciencia y Tecnología (CONACyT), and the Engineering and Physical Sciences Research Council (EPSRC) Ref. EP/M024652/1.

References

[1] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, F. Descamps, A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison, Build. Environ. 75 (2014) 67–78, http://dx.doi.org/10.1016/j.buildenv.2014.01.021.

[2] A. Ballanti, L.F. Ochoa, Initial assessment of voltage-led demand response from UK residential loads, 2015 IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. IEEE (2015) 1–5, http://dx.doi.org/10.1109/ISGT.2015.7131821.

[3] A. Ballanti, F. Pilo, A. Navarro-Espinosa, L.F. Ochoa, Assessing the benefits of PV var absorption on the hosting capacity of LV feeders, IEEE PES ISGT Eur. 2013. IEEE (2013) 1–5, http://dx.doi.org/10.1109/ISGTEurope.2013.6695423.

[4] A. Bondu, A. Dachraoui, Realistic and very fast simulation of individual electricity consumptions, 2015 Int. Jt. Conf. Neural Networks. IEEE (2015) 1–8, http://dx.doi.org/10.1109/IJCNN.2015.7280339.

[5] DECC, Chapter 3: domestic energy consumption, Energy Consum. UK. Department of Energy and Climate Change, London, United Kingdom, Ch. 3 (2015) 17 https://www.gov.uk/government/statistics/energy-consumption-in-the-uk.

[6] E.A. Technology, Assessing the Impact of Low Carbon Technologies on Great Britain's Power Distribution Networks. Tech. Rep., Ofgem, 2012 https://www.ofgem.gov.uk/publications-and-updates/assessing-impact-low-carbon-technologies-great-britains-power-distribution-networks.

[7] Energy Saving Trust, DECC, DEFRA, P. Owen, Powering the Nation: Household Electricity-Using Habits Revealed. Tech. Rep., Energy Saving Trust, Department of Energy and Climate Change, Department for Environment Food and Rural Affairs, London, United Kingdom, 2012 http://www.energysavingtrust.org.uk/Publications2/Corporate/Research-and-insights/Powering-the-nation-household-electricity-using-habits-revealed.

[8] N. Good, L. Zhang, A. Navarro-Espinosa, P. Mancarella, High resolution modelling of multi-energy domestic demand profiles, Appl. Energy 137 (2015) 193–210, http://dx.doi.org/10.1016/j.apenergy.2014.10.028.

[9] Ipsos-RSL, Ipsos-RSL and Office for National Statistics, United Kingdom Time Use Survey, 2003, September, http://dx.doi.org/10.5255/UKDA-SN-4504-1 [computer file]. 3rd Edition. Colchester, Essex: UK Data Archive [distributor], SN: 4504.

[10] E. McKenna, M. Thomson, High-resolution stochastic integrated thermal-electrical domestic demand model, Appl. Energy 165 (2016) 445–461, http://dx.doi.org/10.1016/j.apenergy.2015.12.089.

[11] A. Navarro, L. Ochoa, P. Mancarella, D. Randles, Impacts of photovoltaics on low voltage networks: a case study for the north west of England, in: 22nd Int. Conf. Exhib. Electr. Distrib. (CIRED 2013), Institution of Engineering and Technology, 2013, pp. 1419, http://dx.doi.org/10.1049/cp.2013.1229.

[12] Office for National Statistics, The United Kingdom 2000 Time Use Survey: Technical Report. Tech. Rep., Office for National Statistics, London, United Kingdom, 2003 http://www.ons.gov.uk/ons/rel/lifestyles/time-use/2000-edition/time-use-survey-technical-report.pdf.

[13] Ofgem, Typical Domestic Consumption Values for Gas and Electricity 2015, 2015 https://www.ofgem.gov.uk/gas/retail-market/monitoring-data-and-statistics/typical-domestic-consumption-values.

[14] ONS, Families and Households: 2015. Tech. Rep., Office for National Statistics, London, United Kingdom, 2015 http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2015-11-05.

[15] J. Page, D. Robinson, N. Morel, J.-L. Scartezzini, A generalised stochastic model for the simulation of occupant presence, Energy Build. 40 (2) (2008) 83–98, http://dx.doi.org/10.1016/j.enbuild.2007.01.018.

[16] J. Palmer, N. Terry, T. Kane, S. Firth, M. Hughes, P. Pope, J. Young, D. Knight, D. Godoy-Shimizu, Electrical Appliances at Home: Tuning in to Energy Savings. Tech. Rep., Cambridge Architectural Research – Element Energy – Loughborough University, Cambridge, United Kingdom, 2013 https://www.gov.uk/government/publications/household-electricity-survey-2.

[17] I. Richardson, M. Thomson, D. Infield, A high-resolution domestic building occupancy model for energy demand simulations, Energy Build. 40 (8) (2008) 1560–1566, http://dx.doi.org/10.1016/j.enbuild.2008.02.006.

[18] I. Richardson, M. Thomson, D. Infield, C. Clifford, Domestic electricity use: a high-resolution energy demand model, Energy Build. 42 (10) (2010) 1878–1887, http://dx.doi.org/10.1016/j.enbuild.2010.05.023.

[19] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, Renew. Sustain. Energy Rev. 13 (8) (2009) 1819–1835, http://dx.doi.org/10.1016/j.rser.2008.09.033.

[20] J. Torriti, A review of time use models of residential electricity demand, Renew. Sustain. Energy Rev. 37 (2014) 265–272, http://dx.doi.org/10.1016/j.rser.2014.05.034.

[21] UKERC Energy Data Centre, Electricity Association, Elexon Ltd, Electricity User Load Profiles by Profile Class. Tech. Rep., Electricity Association, London, United Kingdom, 2013 http://data.ukedc.rl.ac.uk/browse/edc/Electricity/LoadProfile/doc.

[22] G. Wahba, Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation, Ann. Stat. 3 (1) (1975) 15–29, http://dx.doi.org/10.1214/aos/1176342997.

[23] J. Widén, A. Molin, K. Ellegård, Models of domestic occupancy, activities and energy use based on time-use data: deterministic and stochastic approaches with application to various building-related simulations, J. Build. Perform. Simul. 5 (1) (2012) 27–44, http://dx.doi.org/10.1080/19401493.2010.532569.

[24] J. Widén, E. Wäckelgård, A high-resolution stochastic model of domestic activity patterns and electricity demand, Appl. Energy 87 (6) (2010) 1880–1892, http://dx.doi.org/10.1016/j.apenergy.2009.11.006.

[25] U. Wilke, F. Haldi, J.-L. Scartezzini, D. Robinson, A bottom-up stochastic model to predict building occupants' time-dependent activities, Build. Environ. 60 (2013) 254–264, http://dx.doi.org/10.1016/j.buildenv.2012.10.021.

[26] R. Yao, K. Steemers, A method of formulating energy load profile for domestic buildings in the UK, Energy Build. 37 (6) (2005) 663–671, http://dx.doi.org/10.1016/j.enbuild.2004.09.007.

[27] J.-P. Zimmermann, M. Evans, J. Griggs, N. King, L. Harding, P. Roberts, C. Evans, Household Electricity Survey: A Study of Domestic Electrical product Usage – Final Report. Tech. Rep., Intertek, London, United Kingdom, 2012 http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&ProjectID=17359&FromSearch=Y&Publisher=1&SearchText=EV0702&SortString=ProjectCode&SortOrder=Asc&Paging=10#Description.